

# TYPICAL DISTANCES IN ULTRASMALL RANDOM NETWORKS

STEFFEN DEREICH, CHRISTIAN MÖNCH, AND PETER MÖRTERS

**Abstract:** We show that in preferential attachment models with power-law exponent  $\tau \in (2, 3)$  the distance between randomly chosen vertices in the giant component is asymptotically equal to  $(4 + o(1)) \frac{\log \log N}{-\log(\tau-2)}$ , where  $N$  denotes the number of nodes. This is twice the value obtained for several types of configuration models with the same power-law exponent. The extra factor reveals the different structure of typical shortest paths in preferential attachment graphs.

*MSc Classification:* Primary 05C80 Secondary 60C05, 90B15.

*Keywords:* Scale-free network, Barabasi-Albert model, preferential attachment, configuration model, dynamical random graph, power law graph, giant component, graph distance, diameter.

## 1. INTRODUCTION

One of the central observations in the theory of scale-free random networks is that in the case of power-law exponents  $\tau \in (2, 3)$  networks are *ultrasmall*, which means that the distance of two randomly chosen nodes in the giant component of a graph with  $N$  vertices is of asymptotic order  $\log \log N$ . The first analytical, but mathematically nonrigorous, evidence for this general phenomenon can be found, for example, in Cohen and Havlin [CH03] or Dorogovtsev et al. [DMS03], and there are also some early papers with rigorous results for specific network models, in particular the work of Reittu and Norros [RN02] and the work of Chung and Lu [CL03].

In the present paper we refine this observation and identify graph distances including constant factors. Our main result is a universal technique for proving lower bounds for typical distances, which in a wide range of examples matches the best upper bounds known from the recent literature. The result is presented in the form of two theorems, which reveal that ultrasmall networks can be divided into two different *universality classes*: For the class of ultrasmall preferential attachment models the typical distances turn out to be twice as large as for models in the class of configuration models. This difference corresponds to different structures of typical shortest paths in the network. We show that the two classes can be easily identified from the form of the attachment probability densities in the networks. We remark here that our work is focused on *typical distances* in networks, as results on diameters tend to be model dependent and universality results are not to be expected.

At least informally, we have some structural insight into typical shortest paths in ultrasmall networks, see for example Norros and Reittu [NR08]. For the class of *configuration models* it turns out that typical vertices in the giant component can be connected with a few steps to a *core* of the network. Within this core there is a hierarchy of *layers* of nodes with increasing connectivity and at the top a small *inner core* of highly connected nodes with very small diameter. A typical shortest path inside the core runs from one layer to the next until the inner core is reached, and then climbing down again until a vertex in the lowest layer of the core is again connected to a typical vertex.

A high degree of a vertex increases its connectivity to any other vertex, and hence the layers can be identified by vertex degrees. Very roughly speaking the  $j$ th layer consists of vertices with degree  $k_j$  where

$$\log k_j \approx (\tau - 2)^{-j}$$

and there are about

$$\frac{\log \log N}{-\log(\tau - 2)}$$

layers. The graph distance of two randomly chosen vertices in the giant component is therefore

$$(2 + o(1)) \frac{\log \log N}{-\log(\tau - 2)}.$$

These asymptotics are rigorously confirmed for two variants of an inhomogeneous random graph model, by Chung and Lu [CL03] and Norros and Reittu [NR06], and for a model with fixed degree sequence by van der Hofstad et al. in [HHZ07]. See also van der Hofstad [Hof10] for a summary of various results with detailed proofs. In general upper bounds on the distances can be obtained by verifying the above strategy, while our Theorem 2 provides a flexible (i.e. model-independent) approach to the lower bound.

For the more complex class of ultrasmall *preferential attachment models* existing results are far less complete. Dommers et al. [DHH10] show that for various ultrasmall preferential attachment models the typical distance of two vertices in the giant component is bounded from above by

$$(4 + o(1)) \frac{\log \log N}{-\log(\tau - 2)}.$$

A corresponding lower bound, and hence confirmation of the exact factor 4, is stated as an interesting open problem by van der Hofstad and Hooghiemstra in [HH08, IV.B] and again in [DHH10], see the remark following Theorem 1.7 and Section 1.2. Our main result, Theorem 1, provides this bound and confirms, somewhat surprisingly, that the upper bound is sharp. Besides the models given in [DHH10] we will also describe other examples of random network models in the same universality class, in which Theorem 1 applies.

Loosely speaking, the shortest paths in the class of preferential attachment models can be described as follows: Again, inside a core of highly connected vertices paths run from bottom to top and back through a hierarchy of layers defined as before. However, by construction of the preferential attachment models a high degree of a vertex does not increase its connectivity to *all* vertices but only to those introduced *late* into the system (which are typically outside the core). Therefore a path cannot directly connect one layer to another in one step, but it requires *two steps*: The paths run from one layer to a young vertex and from there back into the next higher layer. The distance of two typical vertices is therefore increased by a factor of two.

In the following section we formulate the precise results, consisting of two simple hypotheses on a random network leading to the two different lower bound results, see Theorems 1 and 2. The section also contains a brief sketch of the proof technique and introduces the notation used in the proofs. In Section 3 we then discuss several examples of networks in the two universality classes. In all these examples upper bounds can either be found in the literature or derived by simple modifications of these proofs. Section 4 is devoted to the proofs of our main results.

## 2. MAIN RESULTS

A *(dynamic) network model* is a sequence of random graphs  $(\mathcal{G}_N)_{N \in \mathbb{N}}$  with the set of vertices of  $\mathcal{G}_N$  given by  $[N] := \{1, 2, \dots, N\}$  and the set of unoriented edges of  $\mathcal{G}_N$  given by a random symmetric subset of  $[N] \times [N]$ . Occasionally we shall allow multiple edges between the same pair of vertices, but this has no bearing on the connectivity problems discussed here, and is for convenience only. We write  $v \leftrightarrow w$  if the vertices  $v, w$  are connected by an edge in the graph  $\mathcal{G}_N$ . The graph distance is given by

$$d_N(v, w) := \min \{n : \exists v = v_0, v_1, \dots, v_n = w \in \mathcal{G}_N \text{ such that } v_{i-1} \leftrightarrow v_i \ \forall 1 \leq i \leq n\}.$$

The main aim of this paper is to provide techniques to find lower bounds on the *typical distance*, i.e. the asymptotic graph distance of two randomly chosen vertices in the graph  $\mathcal{G}_N$ . Our first result is based on the following assumption.

*Assumption PA( $\gamma$ ):*

There exists  $\kappa$  such that, for all  $N$  and pairwise distinct vertices  $v_0, \dots, v_\ell \in [N]$ ,

$$\mathbb{P}\{v_0 \leftrightarrow v_1 \leftrightarrow v_2 \leftrightarrow \dots \leftrightarrow v_\ell\} \leq \prod_{k=1}^{\ell} \kappa (v_{k-1} \wedge v_k)^{-\gamma} (v_{k-1} \vee v_k)^{\gamma-1}.$$

In preferential attachment models with power law exponent  $\tau$ , Assumption PA( $\gamma$ ) is typically satisfied for all  $\gamma > (\tau - 1)^{-1}$ . Hence we expect these networks to be ultrasmall if and only if  $\frac{1}{2} < \gamma < 1$ . Theorem 1, our main result, gives a lower bound on the typical distance in this case.

**Theorem 1.** *Let  $(\mathcal{G}_N)_{N \in \mathbb{N}}$  be a dynamic network model that satisfies Assumption PA( $\gamma$ ) for some  $\gamma$  satisfying  $\frac{1}{2} < \gamma < 1$ , then, for random vertices  $V$  and  $W$  chosen independently and uniformly from  $[N]$ , we have*

$$d_N(V, W) \geq 4 \frac{\log \log N}{\log(\frac{\gamma}{1-\gamma})} + \mathcal{O}(1)$$

*with high probability.*

Examples of network models, in which Theorem 1 can be applied, will be given as Examples 1–2 in Section 3. They comprise various preferential attachment models with power law exponent  $\tau \in (2, 3)$ . In all these cases Assumption PA( $\gamma$ ) is satisfied for all  $\gamma > (\tau - 1)^{-1}$ , and the theorem implies that

$$d_N(V, W) \geq (4 + o(1)) \frac{\log \log N}{-\log(\tau - 2)}, \quad \text{with high probability as } N \rightarrow \infty.$$

Matching upper bounds are known from the literature.

An approach similar to the above can be used to study lower bounds for the typical distance of ultrasmall *configuration networks*. In this class the connection probabilities look different and we have to formulate a different assumption.

*Assumption* CM( $\gamma$ ):

There exists  $\kappa$  such that, for all  $N$  and pairwise distinct vertices  $v_0, \dots, v_\ell \in [N]$ ,

$$\mathbb{P}\{v_0 \leftrightarrow v_1 \leftrightarrow v_2 \leftrightarrow \dots \leftrightarrow v_\ell\} \leq \prod_{k=1}^{\ell} \kappa v_{k-1}^{-\gamma} v_k^{-\gamma} N^{2\gamma-1}.$$

In configuration models with power law exponent  $\tau$ , Assumption CM( $\gamma$ ) is typically satisfied for all  $\gamma > (\tau - 1)^{-1}$ , and again we expect these networks to be ultrasmall if  $\frac{1}{2} < \gamma < 1$ .

**Theorem 2.** *Let  $(\mathcal{G}_N)_{N \in \mathbb{N}}$  be a dynamic network model that satisfies Assumption CM( $\gamma$ ) for some  $\gamma$  satisfying  $\frac{1}{2} < \gamma < 1$ , then, for random vertices  $V$  and  $W$  chosen independently and uniformly from  $[N]$ , we have*

$$d_N(V, W) \geq 2 \frac{\log \log N}{\log(\frac{\gamma}{1-\gamma})} + \mathcal{O}(1), \quad \text{with high probability as } N \rightarrow \infty.$$

Examples of network models, in which Theorem 2 can be applied, will be given as Examples 3–5 in Section 3. They comprise a variety of configuration models with power law exponent  $\tau \in (2, 3)$ . In all these cases Assumption CM( $\gamma$ ) is satisfied for all  $\gamma > (\tau - 1)^{-1}$ , and the theorem implies that

$$d_N(V, W) \geq (2 + o(1)) \frac{\log \log N}{-\log(\tau - 2)}, \quad \text{with high probability as } N \rightarrow \infty.$$

Again, in all examples matching upper bounds are known from the literature.

The proof of both theorems is based on a *constrained* or *truncated first order method*, which we now briefly explain. We start with an explanation of the (unconstrained) first moment bound and its shortcomings. Let  $v, w$  be distinct vertices of  $\mathcal{G}_N$ . Then, for  $\delta \in \mathbb{N}$ ,

$$\begin{aligned} \mathbb{P}\{d_N(v, w) \leq 2\delta\} &= \mathbb{P}\left(\bigcup_{k=1}^{2\delta} \bigcup_{(v_1, \dots, v_{k-1})} \{v \leftrightarrow v_1 \leftrightarrow \dots \leftrightarrow v_{k-1} \leftrightarrow w\}\right) \\ &\leq \sum_{k=1}^{2\delta} \sum_{(v_1, \dots, v_{k-1})} \prod_{j=1}^k p(v_{j-1}, v_j), \end{aligned}$$

where  $(v_0, \dots, v_k)$  is any collection of pairwise distinct vertices in  $\mathcal{G}_N$  with  $v_0 = v$  and  $v_k = w$  and, for  $m, n \in \mathbb{N}$ ,

$$p(m, n) := \begin{cases} \kappa(m \wedge n)^{-\gamma} (m \vee n)^{\gamma-1} & \text{if PA}(\gamma) \text{ holds;} \\ \kappa m^{-\gamma} n^{-\gamma} N^{2\gamma-1} & \text{if CM}(\gamma) \text{ holds.} \end{cases}$$

Note that one can assign each path  $(v_0, \dots, v_k)$  the weight

$$p(v_0, \dots, v_k) := \prod_{j=1}^k p(v_{j-1}, v_j), \tag{2.1}$$

and the upper bound is just the sum over the weights of all paths from  $v$  to  $w$  of length no more than  $2\delta$ . The shortcoming of this bound is that the paths that contribute most to the total weight are those that connect  $v$ , resp.  $w$ , quickly to vertices with extremely small indices. Since these are typically not present in the network, such paths have to be removed in order to get a reasonable estimate.

To this end we define a decreasing sequence  $(\ell_k)_{k=0,\dots,\delta}$  of positive integers and consider a tuple of vertices  $(v_0, \dots, v_n)$  as *admissible* if  $v_k \wedge v_{n-k} \geq \ell_k$  for all  $k \in \{0, \dots, \delta \wedge n\}$ . We denote by  $A_k^{(v)}$  the event that there exists a path  $v = v_0 \leftrightarrow \dots \leftrightarrow v_k$  in the network such that  $v_0 \geq \ell_0, \dots, v_{k-1} \geq \ell_{k-1}, v_k < \ell_k$ , i.e. a path that traverses the threshold after exactly  $k$  steps. For fixed vertices  $v, w \geq \ell_0$ , the truncated first moment estimate is

$$\mathbb{P}\{d_N(v, w) \leq 2\delta\} \leq \sum_{k=1}^{\delta} \mathbb{P}(A_k^{(v)}) + \sum_{k=1}^{\delta} \mathbb{P}(A_k^{(w)}) + \sum_{n=1}^{2\delta} \sum_{\substack{(v_0, \dots, v_n) \\ \text{admissible}}} \mathbb{P}\{v_0 \leftrightarrow \dots \leftrightarrow v_n\}, \quad (2.2)$$

where the admissible paths in the last sum start with  $v_0 = v$  and end with  $v_n = w$ . By assumption,

$$\mathbb{P}\{v_0 \leftrightarrow \dots \leftrightarrow v_n\} \leq p(v_0, \dots, v_n)$$

so that for  $v \geq \ell_0$  and  $k = 1, \dots, \delta$ ,

$$\mathbb{P}(A_k^{(v)}) \leq \sum_{v_1=\ell_1}^N \dots \sum_{v_{k-1}=\ell_{k-1}}^N \sum_{v_k=1}^{\ell_k-1} p(v, v_1, \dots, v_k). \quad (2.3)$$

Given  $\varepsilon > 0$  we choose  $\ell_0 = \lceil \varepsilon N \rceil$  and  $(\ell_j)_{j=0,\dots,k}$  decreasing fast enough so that the first two summands on the right hand side of (2.2) together are no larger than  $2\varepsilon$ . For  $k \in \{1, \dots, \delta\}$ , set

$$\mu_k^{(v)}(u) := \mathbb{1}_{\{v \geq \ell_0\}} \sum_{v_1=\ell_1}^N \dots \sum_{v_{k-1}=\ell_{k-1}}^N p(v, v_1, \dots, v_{k-1}, u),$$

and set  $\mu_0^{(v)}(u) = \mathbb{1}_{\{v=u\}}$ . To rephrase the truncated moment estimate in terms of  $\mu$ , note that  $p$  is symmetric so that, for all  $n \leq 2\delta$  and  $n^* := \lfloor n/2 \rfloor$ ,

$$\begin{aligned} \sum_{\substack{(v_0, \dots, v_n) \\ \text{admissible}}} \mathbb{P}\{v_0 \leftrightarrow \dots \leftrightarrow v_n\} &\leq \sum_{v_1=\ell_1}^N \dots \sum_{v_{n^*}=\ell_{n^*}}^N \dots \sum_{v_{n-1}=\ell_1}^N p(v, \dots, v_{n^*}) p(v_{n^*}, \dots, w) \\ &= \sum_{v_{n^*}=\ell_{n^*}}^N \mu_{n^*}^{(v)}(v_{n^*}) \mu_{n-n^*}^{(w)}(v_{n^*}). \end{aligned} \quad (2.4)$$

Using the recursive representation

$$\mu_{k+1}^{(v)}(n) = \sum_{m=\ell_k}^N \mu_k^{(v)}(m) p(m, n)$$

we establish upper bounds for  $\mu_k^{(v)}(u)$ , and use these to show that the rightmost term in (2.2) remains small if  $\delta$  is chosen sufficiently small. Using the input from Assumptions PA( $\gamma$ ), resp. CM( $\gamma$ ), this will lead to the lower bounds for the typical distance in both theorems. Detailed proofs will be given in Section 4.

### 3. EXAMPLES

In this section we give five examples, corresponding to the best understood models of ultrasmall networks in the mathematical literature. Examples 1–2 are of preferential attachment type and will be discussed using our main result, Theorem 1, while Examples 3–5 are of configuration type and will be discussed using Theorem 2.

*Example 1* (Preferential attachment with fixed outdegree). This class of models is studied in the work of Hooghiemstra, van der Hofstad and coauthors. We base our discussion on the paper [DHH10], where three qualitatively similar models are considered, see also [Hof10] for a survey. We focus on the first model studied in [DHH10], which is most convenient to define, the two variants can be treated with the same method. The model depends on two parameters, an integer  $\mathfrak{m} \geq 1$  and a real  $\delta > -\mathfrak{m}$ . Roughly speaking, in every step a new vertex is added to the network and connected to  $\mathfrak{m}$  existing vertices with a probability proportional to their degree plus  $\delta$ . Note that in the case  $\mathfrak{m} = 1$  the network has the metric structure of a tree, making this a degenerate case of less interest. The case famously studied by Bollobás and Riordan [BR04] corresponds to  $\delta = 0$  and  $\mathfrak{m} \geq 2$  and leads to a network with  $\tau = 3$  and typical distance  $\log N / \log \log N$ , so that it lies outside the class of ultrasmall networks.

We first generate a dynamic network model  $(\mathcal{G}_N)$  for the case  $\mathfrak{m} = 1$ . By  $Z[n, N]$ ,  $n \leq N$ , we denote the degree of vertex  $n$  in  $\mathcal{G}_N$  (with the convention that self-loops add two towards the degree of the vertex to which they are attached).

- $\mathcal{G}_1$  consists of a single vertex, labelled 1, with one self loop.
- In each further step, given  $\mathcal{G}_N$ , we insert one new vertex, labelled  $N + 1$ , and one new edge into the network such that the new edge connects the new vertex to vertex  $m \in [N]$  with probability

$$\mathbb{P}\{m \leftrightarrow N + 1 \mid \mathcal{G}_N\} = \frac{Z[m, N] + \delta}{N(2 + \delta) + 1 + \delta},$$

or to itself with probability

$$\frac{1 + \delta}{N(2 + \delta) + 1 + \delta}.$$

To generalise the model to arbitrary values of  $\mathfrak{m}$ , we take the graph  $\mathcal{G}'_{\mathfrak{m}N}$  constructed using parameters  $\mathfrak{m}' = 1$  and  $\delta' = \delta/\mathfrak{m}$ , and merge vertices  $\mathfrak{m}(k - 1) + 1, \dots, \mathfrak{m}k$  in the graph  $\mathcal{G}'_{\mathfrak{m}N}$  into a single vertex denoted  $k$ , keeping all edges. We obtain asymptotic degree distributions which are power laws with exponent  $\tau = 3 + \frac{\delta}{\mathfrak{m}}$ , so that we expect to be in the ultrasmall range if and only if  $-\mathfrak{m} < \delta < 0$ .

*Proposition 3.* For independent, uniformly chosen vertices  $V$  and  $W$  in the giant component of the preferential attachment model with parameters  $\mathfrak{m} \geq 2$  and  $-\mathfrak{m} < \delta < 0$ , we have

$$d_N(V, W) = (4 + o(1)) \frac{\log \log N}{-\log(1 + \frac{\delta}{\mathfrak{m}})} \quad \text{with high probability.}$$

*Remark 1.* The upper bound is proved in [DHH10], see the remark following Theorem 1.6. This paper leaves the problem of finding a lower bound open. We resolve this problem by verifying Assumption PA( $\gamma$ ) for  $\gamma = (2 + \frac{\delta}{\mathfrak{m}})^{-1}$  and applying Theorem 1.

*Proof.* We look at  $\mathfrak{m} = 1$  first. In this case, we have, for  $1 \leq m < n \leq N$ ,

$$\mathbb{P}\{m \leftrightarrow n\} = \frac{\mathbb{E}Z[m, n - 1] + \delta}{n(2 + \delta) - 1}. \quad (3.1)$$

It is easy to see that

$$\mathbb{E}[Z[m, n] + \delta \mid Z[m, n - 1]] = (Z[m, n - 1] + \delta) \frac{n(2 + \delta)}{n(2 + \delta) - 1},$$

and hence

$$\mathbb{E}[Z[m, n] + \delta] = (1 + \delta) \frac{\Gamma(n+1)\Gamma(m - \frac{1}{2+\delta})}{\Gamma(n + \frac{1+\delta}{2+\delta})\Gamma(m)}.$$

In particular there exist constants  $0 < c < C$  such that

$$c \left(\frac{n}{m}\right)^{\frac{1}{2+\delta}} \leq \mathbb{E}Z[m, n] \leq C \left(\frac{n}{m}\right)^{\frac{1}{2+\delta}} \quad \text{for all } 1 \leq m < n.$$

Combining this with (3.1) yields, for  $\gamma = \frac{1}{2+\delta}$  and a suitable  $\kappa_1 > 0$ , that

$$\mathbb{P}\{m \leftrightarrow n\} \leq \frac{C(n/m)^\gamma + \delta}{n(2+\delta) - 1} \leq \kappa_1 n^{\gamma-1} m^{-\gamma} \quad \text{for all } 1 \leq m < n. \quad (3.2)$$

To verify  $\text{PA}(\gamma)$ , following [DHH10, Lemma 2.1] we find that for distinct vertices  $v_0, \dots, v_l$  all events of the form  $\{v_{j-1} \leftrightarrow v_j \leftrightarrow v_{j+1}\}$  with  $j \in \{1, \dots, l-1\}$  and  $v_j < v_{j-1}, v_{j+1}$ , and all events  $\{v_{j-1} \leftrightarrow v_j\}$  which are not part of these, are nonpositively correlated, in the sense that the probability of all of them occurring is smaller than the product of the probabilities. Recalling also (3.2) it remains to show that for  $m < v, w$ ,

$$\mathbb{P}\{v \leftrightarrow m \leftrightarrow w\} \leq \kappa_2 v^{\gamma-1} w^{\gamma-1} m^{-2\gamma}, \quad (3.3)$$

for some finite constant  $\kappa_2 > 0$ . To this end we let  $\{(Z_n^{(k,m)})_{n \geq m} : k, m \in \mathbb{N}\}$  denote the collection of right-continuous Markov jump processes starting at  $Z_{m-}^{(k,m)} = k$ , jumping instantly at time  $m$  and subsequently at integer time-steps following the rule

$$\mathbb{P}\{Z_n^{(k,m)} = Z_{n-}^{(k,m)} + 1 \mid Z_{n-}^{(k,m)}\} = \frac{Z_{n-}^{(k,m)} + \delta}{n(2+\delta) - \delta} = 1 - \mathbb{P}\{Z_n^{(k,m)} = Z_{n-}^{(k,m)} \mid Z_{n-}^{(k,m)}\}.$$

Note that  $(Z[m, n])_{n \geq m} = (Z_n^{(1,m)})_{n \geq m}$  in law and that, for  $m < n$ , the event  $\{m \leftrightarrow n\}$  corresponds to  $\{\Delta Z_n^{(k,m)} = 1\}$ , where we write  $\Delta Z_n^{(k,m)} := Z_n^{(k,m)} - Z_{n-}^{(k,m)}$ . Note also that  $Z_n^{(k_0,m)}$  is stochastically dominated by  $Z_n^{(k,m)}$  for  $k \geq k_0$ . Hence, for  $m < n_1 < n_2$ ,

$$\begin{aligned} \mathbb{E}[Z_{n_2}^{(2,m)} \mid \Delta Z_{n_1}^{(2,m)} = 1] &= \sum_{j=2}^{m-n_2+2} \sum_{k=2}^{m-n_1+1} j \mathbb{P}\{Z_{n_2}^{(2,m)} = j \mid Z_{n_1-}^{(2,m)} = k, \Delta Z_{n_1}^{(2,m)} = 1\} \\ &\quad \times \mathbb{P}\{Z_{n_1-}^{(2,m)} = k \mid \Delta Z_{n_1}^{(2,m)} = 1\} \\ &\leq \sum_{j=2}^{m-n_2+2} \sum_{k=2}^{m-n_1+1} \frac{j \mathbb{P}\{Z_{n_2}^{(k+1,n_1)} = j\} (k+\delta) \mathbb{P}\{Z_{n_1-}^{(2,m)} = k\}}{(n_1(2+\delta) + 1 + \delta) \mathbb{P}\{\Delta Z_{n_1}^{(2,m)} = 1\}} \\ &= \sum_{k=2}^{m-n_1+1} \frac{(k+\delta) \mathbb{P}\{Z_{n_1-}^{(2,m)} = k\} \mathbb{E}Z_{n_2}^{(k+1,n_1)}}{(n_1(2+\delta) + 1 + \delta) \mathbb{P}\{\Delta Z_{n_1}^{(2,m)} = 1\}}. \end{aligned}$$

As in the derivation of (3.2) the expectation in the last line can be bounded from above by  $c_0(k+1)n_2^\gamma n_1^{-\gamma}$ , for some  $c_0 > 0$ . Similarly, we obtain  $\mathbb{P}\{\Delta Z_{n_1}^{(2,m)} = 1\} \geq c_1 n_1^{\gamma-1} m^{-\gamma}$  and

$$\mathbb{E}[(Z_{n_1-}^{(2,m)})^2] \leq c_2 m^{-\frac{2}{2+\delta}} n_1^{\frac{2}{2+\delta}},$$

for further constants  $c_1, c_2 > 0$ . Summarising, we obtain

$$\mathbb{E}[Z_{n_2}^{(2,m)} \mid \Delta Z_{n_1}^{(2,m)} = 1] \leq c_3 n_2^\gamma n_1^{-2\gamma} m^\gamma \sum_{k=2}^{m-n_1+1} k^2 \mathbb{P}\{Z_{n_1-}^{(2,m)} = k\} \leq c_4 n_2^\gamma m^{-\gamma},$$

for some  $c_3, c_4 > 0$ , and this establishes (3.3). Finally, passing from  $\mathbf{m} = 1$  to general  $\mathbf{m}$  can be achieved by a simple union bound.  $\square$

A different class of preferential attachment models was introduced in [DM09] and further studied in [DM10]. Here a new vertex is connected to any existing vertex independently with a probability depending (possibly nonlinearly) on its degree. In this model the number of edges created in every step is asymptotically Poisson distributed.

*Example 2* (Preferential attachment with variable outdegree). This model is studied in the work of Dereich, Mörters and coauthors, see [DM11] for a survey. The model depends on a concave function  $f: \mathbb{N} \cup \{0\} \rightarrow (0, \infty)$ , which is called the *attachment rule*. Roughly speaking, in every step a new vertex is added to the network and oriented edges from the new vertex to existing vertices are introduced independently with a probability proportional to the current degree of the existing vertex.

More precisely, to generate a dynamic network model  $(\mathcal{G}_N)$  we assume that  $f$  satisfies  $f(0) \leq 1$  and  $f(1) - f(0) < 1$ . An important parameter derived from  $f$  is the limit

$$\gamma := \lim_{n \rightarrow \infty} \frac{f(n)}{n},$$

which always exists with  $0 \leq \gamma < 1$ , by concavity. By  $Z[n, N]$ ,  $n \leq N$ , we denote the number of *younger* vertices to which vertex  $n$  is connected in  $\mathcal{G}_N$ .

- $\mathcal{G}_1$  consists of a single vertex, labelled 1, and no edges.
- In the  $(N + 1)$ st step, given  $\mathcal{G}_N$ , we insert one new vertex, labelled  $N + 1$ , and independently for any  $m \in [N]$  we introduce an edge from  $N + 1$  to  $m$  with probability

$$\frac{f(Z[m, N])}{N}.$$

By [DM09, Theorem 1.1(b)] the conditional distribution given  $\mathcal{G}_N$  of the number of edges created in the  $(N + 1)$ st step converges to a Poisson distribution and the empirical distribution of the degrees converges to a power law with exponent  $\tau = 1 + \frac{1}{\gamma}$ , or more precisely to a random probability vector  $(\mu_k)$  satisfying

$$\lim_{k \rightarrow \infty} \frac{\log \mu_k}{\log k} = 1 + \frac{1}{\gamma}.$$

We therefore expect the network to be ultrasmall if and only if  $\gamma > \frac{1}{2}$ .

*Proposition 4.* *For independent, uniformly chosen vertices  $V$  and  $W$  in the giant component of the preferential attachment model with attachment rule  $f$  and derived parameter  $\gamma > \frac{1}{2}$ , we have*

$$d_N(V, W) = (4 + o(1)) \frac{\log \log N}{\log(\frac{\gamma}{1-\gamma})} \quad \text{with high probability.}$$

*Remark 2.* The upper bound can be proved by adapting the argument of [DHH10], see the forthcoming thesis [Mön12] for details. For the lower bound we verify Assumption PA( $\gamma + \varepsilon$ ), for any  $\varepsilon > 0$ , and apply Theorem 1.

*Proof.* We first note that, for  $v < w \in [N]$ ,

$$\mathbb{P}\{v \leftrightarrow w\} = \frac{\mathbb{E}f(Z[v, w - 1])}{w - 1}.$$

To estimate the expectation we note that by concavity, given  $\varepsilon > 0$  there exists  $k$  such that, for all  $n \geq k$ , we have  $f(n) \leq f(k) + (\gamma + \varepsilon)(n - k)$ . An easy calculation (see [DM10, Lemma 2.7]) shows that

$$\mathbb{E}f(Z[v, w - 1]) \leq C_1 w^{\gamma + \varepsilon} v^{-\gamma - \varepsilon} \quad \text{for a suitable constant } C_1 > 0. \quad (3.4)$$



We now use (3.4) to verify  $\text{PA}(\gamma + \varepsilon)$ . For  $v < w \in [N]$ , all events  $\{v \leftrightarrow w\}$  with different values of  $v$  are independent. Hence  $\mathbb{P}\{v_0 \leftrightarrow \dots \leftrightarrow v_n\}$  can be decomposed into factors of the form  $\mathbb{P}\{v_{j-1} \leftrightarrow v_j \leftrightarrow v_{j+1}\}$  with  $v_j < v_{j-1}, v_{j+1}$  and factors of the form  $\mathbb{P}\{v_{j-1} \leftrightarrow v_j\}$  for the remaining edges. It remains to estimate factors of the latter form. We may assume  $v < u < w$  and get

$$\mathbb{P}\{u \leftrightarrow v \leftrightarrow w\} = \frac{\mathbb{E}[f(Z[v, u-1])f(Z[v, w-1])]}{(u-1)(w-1)}.$$

Arguing as in the derivation of (3.4) we get, for a suitable constant  $C_2 > 0$ ,

$$\mathbb{E}[f(Z[v, w-1]) \mid Z[v, u-1] = k] \leq C_2 f(k) w^{\gamma+\varepsilon} u^{-\gamma-\varepsilon}.$$

Hence

$$\mathbb{E}[f(Z[v, u-1])f(Z[v, w-1])] \leq C_2 \mathbb{E}[f(Z[v, u-1])^2] w^{\gamma+\varepsilon} u^{-\gamma-\varepsilon},$$

and, using a similar argument as above, we obtain  $C_3 > 0$  such that

$$\mathbb{E}[f(Z[v, u-1])^2] \leq C_3 u^{2\gamma+\varepsilon} v^{-2\gamma-\varepsilon}.$$

Summarising, we obtain a constant  $C_4 > 0$  such that

$$\mathbb{P}\{u \leftrightarrow v \leftrightarrow w\} \leq C_4 u^{\gamma-1+\varepsilon} v^{2\gamma-\varepsilon} w^{\gamma-1+\varepsilon},$$

as required to complete the proof.  $\square$

We now give three examples of random networks in the universality class of configuration models. The first two belong to the wide class of inhomogeneous random graphs, whose essential feature is the independence between different edges.

*Example 3* (Expected degree random graph). This model is studied in the work of Chung and Lu, see [CL03] or [CL06] for a survey. In its general form the model depends on a triangular scheme  $w_1^{(N)}, \dots, w_N^{(N)}$  of positive weights, where the weight  $w_i^{(N)}$  plays the role of the expected degree of vertex  $i$  in  $\mathcal{G}_N$ . The model is defined by the following two requirements:

- for every pair  $(i, j)$  with  $1 \leq i \neq j \leq N$  the events  $\{i \leftrightarrow j\}$  are independent,
- for every pair  $(i, j)$  with  $1 \leq i \neq j \leq N$  we have

$$\mathbb{P}\{i \leftrightarrow j\} = \frac{w_i^{(N)} w_j^{(N)}}{\ell_N} \wedge 1, \quad \text{where } \ell_N := \sum_{i=1}^N w_i^{(N)}.$$

*Proposition 5.* For independent, uniformly chosen vertices  $V$  and  $W$  in the expected degree random graph with weights satisfying

$$c \left(\frac{N}{i}\right)^\gamma \leq w_i^{(N)} \leq C \left(\frac{N}{i}\right)^\gamma \quad \text{for all } 1 \leq i \leq N,$$

for some  $\gamma > \frac{1}{2}$  and constants  $0 < c \leq C$ , we have

$$d_N(V, W) = (2 + o(1)) \frac{\log \log N}{\log(\frac{\gamma}{1-\gamma})} \quad \text{with high probability.}$$

*Proof.* The upper bound is sketched in [CL03]. For the lower bound we have to check Assumption CM( $\gamma$ ). Note that, using the upper bound on the weights,

$$\mathbb{P}\{i \leftrightarrow j\} \leq \frac{w_i^{(N)} w_j^{(N)}}{\ell_N} \leq C^2 \frac{N^{2\gamma}}{\ell_N} (ij)^{-\gamma}.$$

From the lower bound on the weights we get that  $\ell_N \geq cN$ , for some  $c > 0$ , and hence  $\mathbb{P}\{i \leftrightarrow j\} \leq \kappa N^{2\gamma-1} i^{-\gamma} j^{-\gamma}$  for a suitable  $\kappa$ . Using the independence assumption we see that Condition CM( $\gamma$ ) holds, and the lower bound follows from Theorem 2.  $\square$

*Example 4* (Conditionally Poissonian random graph). This model is studied in the work of Norros and Reittu, see [NR06]. It is based on drawing an independent, identically distributed sequence  $\Lambda_1, \Lambda_2, \dots$  of positive capacities. Conditional on this sequence, the dynamical network model is constructed as follows:

- $\mathcal{G}_1$  consists of a single vertex, labelled 1, and no edges.
- In the  $(N + 1)$ st step, given  $\mathcal{G}_N$ , we insert one new vertex, labelled  $N + 1$ , and independently for any  $m \in [N]$  we introduce a random number of edges between  $N + 1$  and  $m$  according to a Poisson distribution with parameter

$$\frac{\Lambda_i \Lambda_{N+1}}{L_{N+1}} \quad \text{for } L_n := \sum_{k=1}^n \Lambda_k.$$

- We further remove each edge in  $\mathcal{G}_N$  independently with probability  $1 - L_N/L_{N+1}$ , and thus obtain  $\mathcal{G}_{N+1}$ .

Recall that having possibly several edges between two vertices has no relevance for the typical distances in the giant component. In order to be in the ultrasmall regime we require the law of the capacities to be power laws with exponent  $2 < \tau < 3$ .

*Proposition 6.* Assume that the capacities in the conditionally Poissonian random graph satisfy

$$\mathbb{P}\{\Lambda_1 > x\} = x^{1-\tau} (c + o(1)) \quad \text{for all sufficiently large } x,$$

where  $2 < \tau < 3$  and  $c > 0$  is constant. For independent, uniformly chosen vertices  $V$  and  $W$  in the giant component we have

$$d_N(V, W) = (2 + o(1)) \frac{\log \log N}{-\log(\tau - 2)} \quad \text{with high probability.}$$

*Remark 3.* The upper bound is proved in [NR06, Theorem 4.2], where it is also shown that a giant component exists. For the lower bound we verify Assumption CM( $\gamma$ ) for  $\gamma = 1/(\tau - 1)$  and apply Theorem 2.

*Proof.* We check that Assumption CM( $\gamma$ ) holds with high probability, conditionally given the capacities. For fixed  $N$  we put the capacities in decreasing order

$$\Lambda_N^{(1)} > \Lambda_N^{(2)} > \dots > \Lambda_N^{(N)}$$

and relabel the vertices so that the  $j$ th vertex has weight  $\Lambda_N^{(j)}$ . We recall from [NR06, Proposition 2.1] that the number of edges between vertices  $i$  and  $j$  in  $\mathcal{G}_N$  is Poisson distributed with parameter  $\Lambda_N^{(i)} \Lambda_N^{(j)} / L_N$ . As the edges are conditionally independent we only have to verify that, given  $\varepsilon > 0$  there exists  $\kappa > 0$  such that

$$1 - \exp\left(-\frac{\Lambda_N^{(i)} \Lambda_N^{(j)}}{L_N}\right) \leq \kappa N^{2\gamma-1} i^{-\gamma} j^{-\gamma} \quad \text{for all } 1 \leq i < j \leq N, \quad (3.5)$$

with probability  $\geq 1 - 2\varepsilon$ . By the law of large numbers  $L_N$  is of order  $N$ , so that it suffices to establish  $\Lambda_N^{(i)} \leq \kappa (N/i)^\gamma$  for all  $1 \leq i \leq N$ . To this end we denote by  $S_N^{(i)}$  the number of potential values exceeding  $\kappa (N/i)^\gamma$ . The random variable  $S_N^{(i)}$  is binomially distributed with parameters  $N$  and  $p := \mathbb{P}\{\Lambda_1 > \kappa (N/i)^\gamma\} \leq c(\kappa) \frac{i}{N}$ , where  $c(\kappa) \downarrow 0$  for  $\kappa \uparrow \infty$ . By Bernstein's inequality, see e.g. [Ben62, (8)],

$$\mathbb{P}\{S_N^{(i)} > 2i\} \leq \exp\left[\frac{-i^2/2}{\text{Var}(S_N^{(i)}) + i/3}\right] \leq e^{-\frac{3}{8}i} \quad \text{if } c(\kappa) < 1.$$

Hence we may choose  $M$  large enough so that  $\sum_{i=M}^\infty \exp(-\frac{3}{8}i) < \varepsilon$ , ensuring that with probability exceeding  $1 - \varepsilon$  we have  $\Lambda_N^{(2i)} \leq \kappa (N/i)^\gamma$  for all  $i \geq M$ . It remains to give bounds

on  $\Lambda_N^{(1)}, \dots, \Lambda_N^{(2M)}$ . By a standard Poisson approximation result, see e.g. [Res08, Proposition 3.21], we note that for any  $1 \leq i \leq 2M$ , we have that  $S_N^{(i)}$  converges weakly to a Poisson distribution with parameter  $\lambda := \lim_{N \rightarrow \infty} N\mathbb{P}\{\Lambda_1 > \kappa(N/i)^\gamma\} \leq 2c(\kappa)M$ , and hence, by choosing  $\kappa$  large, we can ensure that for large  $N$ , we have  $\sum_{i=1}^{2M} \mathbb{P}\{S_N^{(i)} > i\} \leq \varepsilon$ , which completes the proof.  $\square$

A model which also falls in the universality class of configuration models are the random networks with fixed degree sequence<sup>1</sup>. This model is well studied and very detailed results on average distances in the case of power laws with exponent  $\tau \in (2, 3)$  are obtained, in particular by van der Hofstad et al. in [HHZ07].

*Example 5* (Random networks with fixed degree sequence). The idea behind this class of models is to enforce a particular power-law exponent by fixing the degree sequence of the network in a first step. We therefore choose a sequence  $D_1, D_2, \dots$  of independent and identically distributed random variables with values in the nonnegative integers. For given  $N$  we assume that

$$L_N := \sum_{j=1}^N D_j$$

is even, which may be achieved by replacing  $D_N$  by  $D_N - 1$  if necessary. Thus given  $D_1, \dots, D_N$  we construct the network  $\mathcal{G}_N$  as follows:

- To any vertex  $m \in [N]$  we attach  $D_m$  half-edges or stubs.
- The  $L_N$  stubs are given an (arbitrary) order.
- We start by pairing the first stub with a (uniformly) randomly chosen other stub, and continue pairing the lowest numbered unpaired stub with a remaining randomly chosen stub until all stubs are matched.
- Any pair of stubs are connect to form an edge.

Obviously the resulting network can have self-loop and double edges, but this has no relevance for the typical distances in the giant component. In order to be in the ultrasmall regime we require the law of the degrees to be a power law with exponent  $2 < \tau < 3$ .

*Proposition 7.* Assume that there exists  $c > 0$  such that

$$\mathbb{P}\{D_1 > x\} = x^{1-\tau} (c + o(1)) \quad \text{for all sufficiently large } x.$$

For independent, uniformly chosen vertices  $V$  and  $W$  in the giant component we have

$$d_N(V, W) = (2 + o(1)) \frac{\log \log N}{-\log(\tau - 2)} \quad \text{with high probability.}$$

*Remark 4.* This and much more is proved in [HHZ07, Theorem 1.2]. For an alternative approach to the lower bound we now verify Assumption CM( $\gamma$ ) for any  $\gamma < 1/(\tau - 1)$  and paths of length up to  $\ell = \mathcal{O}(\log \log N)$ , which is clearly sufficient to apply Theorem 2.

*Proof.* We observe that, given  $D_1, \dots, D_N$ , for pairwise disjoint vertices  $v_1, \dots, v_\ell, v_{\ell+1}$ ,

$$\mathbb{P}\{v_\ell \leftrightarrow v_{\ell+1} \mid v_1 \leftrightarrow v_2 \leftrightarrow \dots \leftrightarrow v_{\ell-1} \leftrightarrow v_\ell\} \leq \frac{D_{v_\ell} D_{v_{\ell+1}}}{L_N - 2 \sum_{k=1}^{\ell} D_{v_k}},$$

---

<sup>1</sup>In fact, in the literature these models are often called configuration models. We prefer to use the term for the wider class of models where vertices are equipped with an a-priori configuration of individual features.

where the denominator is a rough lower bound on the number of stubs unaffected by the conditioning event. In particular,  $\mathbb{P}\{i \leftrightarrow j\} \leq \frac{D_i D_j}{L_N - 2D_i}$ . Using the law of large numbers one can easily see that there is a  $c > 0$  such that

$$L_N - 2 \sum_{k=1}^{\ell} D_{v_k} \geq cN \quad \text{with high probability,}$$

for any choice of  $v_1, \dots, v_{\ell}$ , if  $\ell = \mathcal{O}(\log \log N)$ . Therefore, to verify Assumption CM( $\gamma$ ) we only need to find appropriate bounds on the degrees of given vertices, which can be achieved (using the same relabeling) by a similar argument as in Example 4.  $\square$

#### 4. PROOFS

**4.1. Proof of Theorem 1.** In this section, we assume validity of Assumption PA( $\gamma$ ) for a  $\gamma \in (\frac{1}{2}, 1)$  with a fixed constant  $\kappa$ . Given a vector  $(q(1), \dots, q(n))$  we use the notation

$$q[m] := \sum_{i=1}^m q(i) \quad \text{for all } 1 \leq m \leq n.$$

We adopt the notation of the discussion at the end of Section 2. In particular recall the definition of  $\mu_k^{(v)}$  and the key estimates (2.2), (2.3) and (2.4), which combined give

$$\mathbb{P}\{d_N(v, w) \leq 2\delta\} \leq \sum_{k=1}^{\delta} \mu_k^{(v)}[\ell_k - 1] + \sum_{k=1}^{\delta} \mu_k^{(w)}[\ell_k - 1] + \sum_{n=1}^{2\delta} \sum_{u=\ell_{n^*}}^N \mu_{n^*}^{(v)}(u) \mu_{n-n^*}^{(w)}(u). \quad (4.1)$$

The remaining task of the proof is to choose  $\delta \in \mathbb{N}$  and  $2 \leq \ell_{\delta} \leq \dots \leq \ell_0 \leq N$  which allow the required estimates for the right hand side. To do so we will make use of the recursive representation

$$\mu_{k+1}^{(v)}(n) = \sum_{m=\ell_k}^N \mu_k^{(v)}(m) p(m, n) \quad \text{for } k \in \{0, \dots, \delta - 1\} \text{ and } n \in [N],$$

where  $\mu_0^{(v)}(n) = \mathbb{1}\{v = n\}$  and

$$p(m, n) = \kappa(m \wedge n)^{-\gamma} (m \vee n)^{\gamma-1}.$$

Denote by  $\bar{\mu}_k^{(v)}(m) = \mathbb{1}_{\{m \geq \ell_k\}} \mu_k^{(v)}(m)$  the truncated version of  $\mu_k^{(v)}$  and conceive  $\mu_k^{(v)}$  and  $\bar{\mu}_k^{(v)}$  as row vectors. Then

$$\mu_{k+1}^{(v)} = \bar{\mu}_k^{(v)} \mathbf{P}_N, \quad (4.2)$$

where  $\mathbf{P}_N = (p(m, n))_{m, n=1, \dots, N}$ . Our aim is to provide a majorant of the form

$$\mu_k^{(v)}(m) \leq \alpha_k m^{-\gamma} + \mathbb{1}_{\{m > \ell_{k-1}\}} \beta_k m^{\gamma-1} \quad (4.3)$$

for suitably chosen parameters  $\alpha_k, \beta_k \geq 0$ . Key to this choice is the following lemma.

**Lemma 8.** *Suppose that  $2 \leq \ell \leq N$ ,  $\alpha, \beta \geq 0$  and  $q: [N] \rightarrow [0, \infty)$  satisfies*

$$q(m) \leq \mathbb{1}\{m \geq \ell\} (\alpha m^{-\gamma} + \beta m^{\gamma-1}) \quad \text{for all } m \in [N].$$

*Then there exists a constant  $c > 1$  (depending only on  $\gamma$  and  $\kappa$ ) such that*

$$q \mathbf{P}_N(m) \leq c \left( \alpha \log\left(\frac{N}{\ell}\right) + \beta N^{2\gamma-1} \right) m^{-\gamma} + \mathbb{1}\{m > \ell\} c \left( \alpha \ell^{1-2\gamma} + \beta \log\left(\frac{N}{\ell}\right) \right) m^{\gamma-1}$$

*for all  $m \in [N]$ .*

*Proof.* One has

$$\begin{aligned}
q\mathbf{P}_N(m) &= \mathbb{1}\{m > \ell\} \sum_{k=\ell}^{m-1} q(k) p(k, m) + \sum_{k=m \vee \ell}^N q(k) p(k, m) \\
&\leq \mathbb{1}\{m > \ell\} \sum_{k=\ell}^{m-1} \kappa(\alpha k^{-\gamma} + \beta k^{\gamma-1}) k^{-\gamma} m^{\gamma-1} + \sum_{k=m \vee \ell}^N \kappa(\alpha k^{-\gamma} + \beta k^{\gamma-1}) k^{\gamma-1} m^{-\gamma} \\
&\leq \kappa \left( \alpha \sum_{k=m \vee \ell}^N k^{-1} + \beta \sum_{k=m \vee \ell}^N k^{2\gamma-2} \right) m^{-\gamma} \\
&\quad + \mathbb{1}\{m > \ell\} \kappa \left( \alpha \sum_{k=\ell}^{m-1} k^{-2\gamma} + \beta \sum_{k=\ell}^{m-1} k^{-1} \right) m^{\gamma-1} \\
&\leq \kappa \left( \alpha \log \left( \frac{m}{\ell-1} \right) + \frac{\beta}{2\gamma-1} N^{2\gamma-1} \right) m^{-\gamma} \\
&\quad + \mathbb{1}\{m > \ell\} \kappa \left( \frac{\alpha}{1-2\gamma} (\ell-1)^{1-2\gamma} + \beta \log \left( \frac{m}{\ell-1} \right) \right) m^{\gamma-1}.
\end{aligned}$$

This implies immediately the assertion since  $\ell \geq 2$  by assumption.  $\square$

We apply Lemma 8 iteratively. Fix  $\varepsilon > 0$  small and start with

$$\ell_0 = \lceil \varepsilon N \rceil, \alpha_1 = \kappa(\varepsilon N)^{\gamma-1} \text{ and } \beta_1 = \kappa(\varepsilon N)^{-\gamma}.$$

Fix  $v \geq \ell_0$ . Then, for all  $m \in [N]$ ,

$$\begin{aligned}
\mu_1^{(v)}(m) &= p(v, m) \leq \kappa \ell_0^{\gamma-1} m^{-\gamma} + \mathbb{1}\{m > \ell_0\} \kappa \ell_0^{-\gamma} m^{\gamma-1} \\
&\leq \alpha_1 m^{-\gamma} + \mathbb{1}\{m > \ell_0\} \beta_1 m^{\gamma-1}.
\end{aligned}$$

Now suppose, for some  $k \in \mathbb{N}$ , we have chosen  $\alpha_k, \beta_k$  and an integer  $\ell_{k-1}$  such that

$$\mu_k^{(v)}(m) \leq \alpha_k m^{-\gamma} + \beta_k m^{\gamma-1} \text{ for all } m \in [N].$$

We choose  $\ell_k$  as an integer satisfying

$$\frac{6\varepsilon}{\pi^2 k^2} \geq \frac{1}{1-\gamma} \alpha_k \ell_k^{1-\gamma}, \quad (4.4)$$

and assume  $\ell_k \geq 2$ . Pick  $\alpha_k, \beta_k$  such that

$$\begin{aligned}
\alpha_{k+1} &\geq c \left( \alpha_k \log \left( \frac{N}{\ell_k} \right) + \beta_k N^{2\gamma-1} \right), \\
\beta_{k+1} &\geq c \left( \alpha_k \ell_k^{1-2\gamma} + \beta_k \log \left( \frac{N}{\ell_k} \right) \right).
\end{aligned} \quad (4.5)$$

By the induction hypothesis we can apply Lemma 8 with  $\ell = \ell_k$  and  $q(m) = \bar{\mu}_k^{(v)}(m)$ . Then, using (4.2),

$$\mu_{k+1}^{(v)}(m) \leq \alpha_{k+1} m^{-\gamma} + \mathbb{1}\{m > \ell_k\} \beta_{k+1} m^{\gamma-1} \text{ for all } m \in [N], \quad (4.6)$$

showing that the induction can be carried forward up to the point where  $\ell_k < 2$ .

Summing over (4.6) and using (4.4) we obtain

$$\mu_k^{(v)}[\ell_k - 1] \leq \frac{1}{1-\gamma} \alpha_k \ell_k^{1-\gamma} \leq \frac{6\varepsilon}{\pi^2 k^2}.$$

Hence the first two summands on the right hand side in (4.1) are together smaller than  $2\varepsilon$ . It remains to choose  $\delta = \delta(N)$  as large as possible while ensuring that  $\ell_\delta \geq 2$  and

$$\lim_{N \rightarrow \infty} \sum_{n=1}^{2\delta} \sum_{u=\ell_{n^*}}^N \mu_{n^*}^{(v)}(u) \mu_{n-n^*}^{(w)}(u) = 0.$$

To this end assume that  $\ell_k$  is the largest integer satisfying (4.4) and the parameters  $\alpha_k, \beta_k$  are defined via equalities in (4.5). To establish lower bounds for the decay of  $\ell_k$  we investigate the growth of  $\eta_k := N/\ell_k > 0$ . Going backwards through the definitions yields, for  $k \geq 1$ , that

$$\left(\eta_{k+2}^{-1} + \frac{1}{N}\right)^{\gamma-1} \leq \frac{c^2(k+2)^2}{k^2} \eta_k^\gamma + 2c \frac{(k+2)^2}{(k+1)^2} \eta_{k+1}^{1-\gamma} \log \eta_{k+1},$$

with  $\eta_1, \eta_2 \leq C_0$  for some constant  $C_0 > 0$  (which, as all constants in this paragraph, may depend on  $\varepsilon$ ). It is easy to check inductively that for any solution of this system there exist constants  $b, B > 0$  such that,

$$\eta_k \leq b \exp\left(B\left(\sqrt{\frac{\gamma}{1-\gamma}}\right)^k\right). \quad (4.7)$$

We now use (4.6) to estimate

$$\begin{aligned} \sum_{n=1}^{2\delta} \sum_{u=\ell_k}^N \mu_{n^*}^{(v)}(u) \mu_{n-n^*}^{(w)}(u) &\leq 2 \sum_{k=1}^{\delta} \sum_{u=\ell_k}^N (\alpha_k u^{-\gamma} + \beta_k u^{\gamma-1})^2 \\ &\leq \frac{4}{2\gamma-1} \sum_{k=1}^{\delta} (\alpha_k^2 \ell_k^{1-2\gamma} + \beta_k^2 N^{2\gamma-1}) \leq \frac{4}{2\gamma-1} \delta (\alpha_\delta^2 \ell_\delta^{1-2\gamma} + \beta_\delta^2 N^{2\gamma-1}). \end{aligned}$$

Using (4.4) and (4.7) the first summand in the bracket can be estimated by

$$\alpha_\delta^2 \ell_\delta^{1-2\gamma} \leq (\delta^{-2} \frac{6\varepsilon}{\pi^2} (1-\gamma))^2 \ell_\delta^{-1} \leq \left(\frac{6\varepsilon}{b\pi^2} (1-\gamma)\right)^2 \frac{1}{N\delta^4} \exp\left(B\left(\frac{\gamma}{1-\gamma}\right)^{\delta/2}\right).$$

Using equality in (4.5) we get  $\beta_\delta \leq c(\alpha_\delta \ell_\delta^{1-2\gamma} + \alpha_\delta N^{1-2\gamma} \log(N/\ell_\delta))$ . Noting that the second summand on the right hand side is bounded by a multiple of the first, we find a constant  $C_1 > 0$  such that  $\beta_\delta^2 N^{2\gamma-1} \leq C_1 \alpha_\delta^2 \ell_\delta^{1-2\gamma}$ , and thus, for a suitable constant  $C_2 > 0$ ,

$$\sum_{n=1}^{2\delta} \sum_{u=\ell_k}^N \mu_{n^*}^{(v)}(u) \mu_{n-n^*}^{(w)}(u) \leq C_2 \frac{1}{N\delta^3} \exp\left(B\left(\frac{\gamma}{1-\gamma}\right)^{\delta/2}\right).$$

Hence, for a suitable constant  $C > 0$ , choosing

$$\delta \leq \frac{\log \log N}{\log \sqrt{\frac{\gamma}{1-\gamma}}} - C$$

we obtain that the term we consider goes to zero of order  $\mathcal{O}((\log \log N)^{-3})$ . Note from (4.7) that this choice also ensures that  $\ell_\delta \geq 2$ . We have thus shown that

$$\mathbb{P}\{d_N(v, w) \geq 2\delta\} \leq 2\varepsilon + \mathcal{O}((\log \log N)^{-3}),$$

whenever  $v, w \geq \ell_0 = \lceil \varepsilon N \rceil$ , which implies the statement of Theorem 1.

**4.2. Proof of Theorem 2.** In this section, we assume validity of Assumption CM( $\gamma$ ) for some  $\gamma \in (\frac{1}{2}, 1)$  with a fixed constant  $\kappa \geq 1$ . Recall again the notation and framework from the introductory chapter. We use the same approach as in the proof of Theorem 1 but now we have to consider the matrix  $\mathbf{P}_N := (p(m, n))_{m, n \in [N]}$  given by

$$p(m, n) := \kappa m^{-\gamma} n^{-\gamma} N^{2\gamma-1} \text{ for } m, n \in [N]. \quad (4.8)$$

We obtain the following lemma, which is the analogue of Lemma 8.

**Lemma 9.** *Suppose that  $2 \leq \ell \leq N$  and  $q: [N] \rightarrow [0, \infty)$  satisfies*

$$q(m) \leq \mathbb{1}\{m \geq \ell\} m^{\gamma-1} \ell^{-\gamma} \text{ for all } m \in [N].$$

*Then, for all  $m \in [N]$ ,*

$$q\mathbf{P}_N(m) \leq \kappa m^{-\gamma} N^{\gamma-1} \left(\frac{N}{\ell}\right)^\gamma \log\left(\frac{N-1}{\ell-1}\right).$$

*Proof.* By (4.8) and the assumption on  $q$ ,

$$q\mathbf{P}_N(m) = \sum_{i=1}^N q(i)p(i, m) \leq \kappa m^{-\gamma} \ell^{-\gamma} N^{2\gamma-1} \sum_{i=\ell}^N \frac{1}{i} \leq \kappa m^{-\gamma} \ell^{-\gamma} N^{2\gamma-1} \log\left(\frac{N-1}{\ell-1}\right),$$

which implies the statement of the lemma.  $\square$

For fixed  $\varepsilon > 0$  we first construct inductively a strictly decreasing sequence of integers  $(\ell_k)_{k=0, \dots, \delta}$  by letting  $\ell_0 = \lceil \varepsilon N \rceil$  and defining  $\ell_{k+1}$  as the largest integer such that, given  $\ell_k \geq 2$ ,

$$\frac{\kappa}{1-\gamma} \left(\frac{\ell_{k+1}}{N}\right)^{1-\gamma} \leq \frac{6\varepsilon}{\pi^2(k+1)^2} \left(\log\left(\frac{N-1}{\ell_k-1}\right)\right)^{-1} \left(\frac{\ell_k}{N}\right)^\gamma. \quad (4.9)$$

Recall the definition and recursive formula for  $\mu_k^{(v)}$  and let  $\bar{\mu}_k^{(v)}(m) := \mathbb{1}\{m \geq \ell_k\} \mu_k^{(v)}(m)$ . Then  $\mu_{k+1}^{(v)}(m) = \bar{\mu}_k^{(v)} \mathbf{P}_N(m)$ . We now apply inductively Lemma 9 and obtain,

$$\mu_k^{(v)}(m) \leq \kappa m^{-\gamma} N^{\gamma-1} \left(\frac{N}{\ell_{k-1}}\right)^\gamma \log\left(\frac{N-1}{\ell_{k-1}-1}\right) \leq m^{-\gamma} \ell_k^{\gamma-1}, \text{ for all } m \in [N]. \quad (4.10)$$

Note that the second inequality in (4.10) follows from (4.9), and hence  $\bar{\mu}_k^{(v)}(m) \leq m^{\gamma-1} \ell_k^{-\gamma}$ , which allows us to continue the induction. Considering the truncated first moment estimate (2.2) for our choice of  $(\ell_k)_{k=0, \dots, \delta}$ , we obtain from (4.10) that

$$\mathbb{P}(A_k^{(v)}) \leq \mu_k^{(v)}[\ell_k - 1] \leq \frac{\kappa}{1-\gamma} \left(\frac{\ell_k}{N}\right)^{1-\gamma} \left(\frac{N}{\ell_{k-1}}\right)^\gamma \log\left(\frac{N-1}{\ell_{k-1}-1}\right).$$

Hence (4.9) entails that  $\sum_{k=1}^{\delta} \mathbb{P}(A_k^{(v)}) \leq \varepsilon$ . The last step is to choose  $\delta = \delta(N)$  as large as possible while ensuring that  $\ell_\delta \geq 2$  and

$$\lim_{N \rightarrow \infty} \sum_{n=1}^{2\delta} \sum_{u=\ell_{n^*}}^N \mu_{n^*}^{(v)}(u) \mu_{n-n^*}^{(w)}(u) = 0. \quad (4.11)$$

By (4.10) the term on the left can be bounded by a constant multiple of  $N^{2\gamma-2} \sum_{k=1}^{\delta} \ell_k^{1-2\gamma}$ . To verify (4.11) we have to bound the growth of the values  $\eta_k := \frac{N}{\ell_k}$ . The choice made in (4.9) implies that  $(\eta_k)_{k \geq 0}$  obeys  $\eta_0 \leq \varepsilon^{-1}$  and

$$\left(\eta_{k+1}^{-1} + \frac{1}{N}\right)^{\gamma-1} < \frac{\pi^2 \kappa}{1-\gamma} \frac{(k+1)^2}{6\varepsilon} \eta_k^\gamma \log(2\eta_k), \text{ for } k \geq 0.$$

From this it is straightforward to verify inductively the existence of constants  $b, B > 0$ , which only depend on  $\varepsilon, \kappa$  and  $\gamma$ , such that

$$\eta_k \leq b \exp \left( B \left( \frac{\gamma}{1-\gamma} \right)^k \right), \text{ for } k \geq 0.$$

Hence, we may choose a suitable constant  $C > 0$  such that for

$$\delta \leq \frac{\log \log N}{\log \left( \frac{\gamma}{1-\gamma} \right)} - C$$

we have  $\ell_\delta \geq 2$ . To complete the proof, we note that

$$N^{2\gamma-2} \sum_{k=1}^{\delta} \ell_k^{1-2\gamma} \leq \frac{1}{N} \sum_{k=1}^{\delta} \eta_k^{2\gamma-1} \leq \delta b N^{B \left( \frac{\gamma}{1-\gamma} \right)^{-C} - 1},$$

which implies convergence in (4.11) when  $C$  is chosen large enough.

**Acknowledgements:** The third author would like to acknowledge the support of *EPSRC* through the award of an *Advanced Research Fellowship*.

## REFERENCES

- [Ben62] G. Bennett. Probability inequalities for the sum of independent random variables. *Journal of the American Statistical Association*, 57:33–45, 1962.
- [BR04] B. Bollobás and O. Riordan. The diameter of a scale-free random graph. *Combinatorica*, 24:5–34, 2004.
- [CH03] R. Cohen and S. Havlin. Scale-free networks are ultrasmall. *Physical Review Letters*, 90:058701, 2003.
- [CL03] F. Chung and L. Lu. The average distance in a random graph with given expected degrees. *Internet Mathematics*, 1:91–113, 2003.
- [CL06] F. Chung and L. Lu. *Complex graphs and networks*, volume 107 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC, 2006.
- [DHH10] S. Dommers, R. van der Hofstad, and G. Hooghiemstra. Diameters in preferential attachment models. *Journal of Statistical Physics*, 139:72–107, 2010.
- [DM09] S. Dereich and P. Mörters. Random networks with sublinear preferential attachment: degree evolutions. *Electronic Journal of Probability*, 14:1222–1267, 2009.
- [DM10] S. Dereich and P. Mörters. Random networks with sublinear preferential attachment: the giant component. *Preprint arXiv:1007.0899. Submitted for publication*, 2010.
- [DM11] S. Dereich and P. Mörters. Random networks with concave preferential attachment rule. *Jahresbericht der Deutschen Mathematiker Vereinigung*, 2011.
- [DMS03] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin. Metric structure of random networks. *Nuclear Physics B*, 653:307–338, 2003.
- [HH08] R. van der Hofstad and G. Hooghiemstra. Universality for distances in power-law random graphs. *Journal of Mathematical Physics*, 49:125209, 2008.
- [HHZ07] R. van der Hofstad, G. Hooghiemstra, and D. Znamenski. Distances in random graphs with finite mean and infinite variance degrees. *Electronic Journal of Probability*, 12:703–766, 2007.
- [Hof10] R. van der Hofstad. Random graphs and complex networks. 2010.
- [Mön12] C. Mönch. *Distances in preferential attachment networks*. PhD thesis, University of Bath, 2012.
- [NR06] I. Norros and H. Reittu. On a conditionally Poissonian graph process. *Advances in Applied Probability*, 38:59–75, 2006.
- [NR08] I. Norros and H. Reittu. Network models with a ‘soft hierarchy’: A random graph construction with loglog scalability. *IEEE Network*, 22:40–46, 2008.
- [Res08] S. I. Resnick. *Extreme values, regular variation, and point processes*. Springer, New York, 2008.
- [RN02] H. Reittu and I. Norros. On the effect of very large nodes in internet graphs. In *Globecom’02*, Vol. III (Proc. Global Telecommunications Conf., Taipei, 2002), pages 2624–2628, 2002.